A Summary of Previous Grand Challenge Proposals for Cognitive Systems

Samuel Bayer
Laurie Damianos
Lynette Hirschman
Gary Strong

The MITRE Corporation

Prepared for DARPA IPTO, September 2004

Version 1.4

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 2004		**00-09-2004 to 00-09-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Summary of Previous Grand Challenge Proposals for Cognitive Systems	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **MITRE Corporation,202 Burlington Road,Bedford,MA,01730-1420**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

1 Overview

The notion of a Grand Challenge (GC) in computational cognition is not new. It has been addressed both specifically and in the context of Grand Challenges in computing as a whole. One well-known example, DARPA's Autonomous Vehicle Grand Challenge (AVGC), has captured the imagination of the media and the public. The AVGC is much more than a compelling research goal or a way to make DARPA's work relevant to the average layperson; it is a measurable test which can tell us where to focus our work and how much we have accomplished. The AVGC has “raised the bar” for what it means for a Grand Challenge to set the agenda for a field of research.

There have been previous efforts to develop Grand Challenges for computer science, but none of these efforts has addressed directly the needs of DARPA IPTO, in particular, demonstrations of cognitive capabilities with a dimension in learning.

To gain insight into why no proposal has yet to become an IPTO Grand Challenge, we performed a historical review and analysis of several sources of GCs in cognitive systems and artificial intelligence (Appendix B). This document summarizes and characterizes these previous Grand Challenge explorations and evaluates categories of proposals against the DARPA IPTO criteria for selecting a GC.

2 Criteria for Selecting an IPTO Grand Challenge

We compiled relevant criteria for selecting a GC from the sources listed in Appendix B, with respect to IPTO-specific requirements. IPTO further refined the compilation, resulting in the following six criteria, with specific components, for selecting an IPTO Grand Challenge.

1. Clear and compelling demonstration of cognition.
    a. The test should be a proxy for a range of problems requiring cognitive capabilities.
    b. The test should not be "game-able" or solvable by "cheap tricks"
    c. It should not be solvable by brute force computation, alone, and it should not lend itself to idiot savant solutions.
    d. Require integration of multiple cognitive capabilities.
        i. It is desirable that the portfolio of tests include sensing and acting (i.e., situated cognition)
2. Clear and simple measurement.
    a. The test should have a clear and simple method for measuring success.
    b. The test should specify what must be done, not how to do it.
    c. It is desirable to have a graduated sequence of increasingly more difficult problems.
    d. It is desirable to have tests that are automatically score-able.
    e. It is desirable that the tests be easy to create and run and that test results be reproducible.
3. Decomposable and diagnostic.
    a. The test should be decomposable into sub-tests or sub-measurements for different aspects of cognition.
    b. The test should be diagnostic (failure to pass the test should point the way to future improvements).
    c. It would be desirable to have partial, intermediate results (scores are not just "Pass/Fail."
4. Ambitious and visionary, but not unrealistic.
    a. It should not be a toy problem.
    b. It should represent technical/scientific goals achievable within a 10-20 year window.
    c. It should not be something that a computer can already do.
    d. Desirable to have military relevance (eventual)
5. Compelling to the general public.
    a. It should be simple to explain and convey to the general public.
6. Motivating for the researchers.
    a. It should generate enthusiasm in the research community.
    b. It is desirable to have a low cost of entry so that work on the problem can begin right away.
    c. It is desirable to enable continuous testing, perhaps over the web.

3 A Review of Previous Grand Challenges

For historical purposes, we collected, compiled, and reviewed many proposed Grand Challenges (see Appendix A for a brief listing). In general, we found that proposals focusing on problems without specifying details of the solution do not provide enough direction for a GC. For example, "Use computational cognition to solve the problem of unemployment." Alternately, proposals focusing on specific cognitive capabilities, without specifying how those capabilities will be used, (e.g., "Learn to Speak as Well as a Human") are difficult to measure.

We chose to focus our analysis on task-based GCs as the most appropriate for IPTO. Task-based GCs are more likely to be organized around a goal whose achievement can be measured, decomposable and diagnostic, and whose usefulness and relevance is clear. An example of one such task-based GCs is "Lead an Orienteering Team to Victory."

For purposes of discussion, we have clustered all GC proposals into categories. (Note that some proposals may be grouped incorrectly due to lack of detail.) We then evaluated each proposal against the criteria for selecting an IPTO GC and summarized these evaluations, by category, in Table 1.

Most of these criteria do not lend themselves in all cases to a yes or no answer. In our evaluation, we used a '+' sign to indicate that a category rated highly against a criterion for all or most GCs in that category and a '-' where the category rated poorly against a criterion. Where different GCs within a single category rated differently, or where ratings were ambiguous, we used no marking at all. Unknown values are indicated by a '?.'

The results of our evaluation indicate that no single GC category is strong in all areas of the criteria that are important to IPTO. While it is difficult to judge whether a GC will be motivating to researchers (6a) or simple to explain (5a), it seems that GCs fail more often than not to be clear and simple to measure (2) or decomposable and diagnostic (3).

Criteria	Grand Challenge Categories														
	Take a Test	Analyze and Persuade	Learn Then Do / Learn Then Teach	Play a Game	Location-Aware Logistical Support	Personal Assistant	Scientific Support	Communication Support	Physical Activities	Collaboration Support	Creative Activities	Question Answering	Prediction	Human Impersonation	Deception Detection
1. Clear & compelling demonstration of cognition															
a. Proxy for problems requiring cognitive capabilities	+	+	+	+	+	+		+	+	+	+	+		+	+
b. Not "game-able" or solvable by "cheap tricks"	+	+	+	+	+	+		+	+	+	+	+		+	+
c. Not be solvable by brute force or idiot savant solutions	+	+	+	+	+	+		+	+	+	+	+		+	+
d. Multiple cognitive capabilities	+	+	+	+	+	+		+	+	+	+	+		+	+
2. Clear & simple measurement															
a. Clear & simple measure of success	+	-		+	-	-	-			-	-		+		
b. Specify what, not how	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
c. Sequence of increasingly difficult problems	+		+		+			+	+	+		+	-	-	+
d. Automatically score-able	+	-			-	-	-	-		-	-		+	-	+
e. Tests easy to run & reproducible results	+		+		-	+			+				-	-	+
3. Decomposable & diagnostic															
a. Decomposable into sub-tests or sub-measurements		-		-					+				-	-	
b. Diagnostic				+		-	-		+		-		-	-	-
c. Intermediate results	+			+	+			+	+	+	+	+	-		-

Table 1 Grand Challenge Proposal Categories Rated Against the IPTO Criteria (continued on next page…)

(continued)

Criteria	Grand Challenge Categories														
	Take a Test	Analyze and Persuade	Learn Then Do / Learn Then Teach	Play a Game	Location-Aware Logistical Support	Personal Assistant	Scientific Support	Communication Support	Physical Activities	Collaboration Support	Creative Activities	Question Answering	Prediction	Human Impersonation	Deception Detection
4. Ambitious & visionary, not unrealistic															
a. Not toy problem	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
b. Goals within 10-20 year window	+	+	+	+	+	+	+	+	+	+	+	+		-	+
c. Not do-able now	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
d. Military relevance			+		+	+		+	+	+		+		+	+
5. Compelling to public															
a. Simple to explain	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
6. Motivating for researchers															
a. Generate enthusiasm	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
b. Low cost of entry	+	+	+	+	+	+	+	+		+	+	+	+	+	+
c. Continuous testing		-	-	+	-	+				-		-	+		-

Table 1 Grand Challenge Proposal Categories Rated Against the IPTO Criteria. + means a GC category ranks highly with respect to a specific criterion, - means a category fails to meet the criterion, ? means unknown, and blank values indicate an ambiguous rating or both positive and negative ratings within the same category.

A. Appendix: Categorized Grand Challenge Proposals

This table represents one of many possible clusterings of Grand Challenge proposals. Note that some proposals may be incorrectly categorized due to lack of detail.

Grand Challenge Categories	Grand Challenge Proposals	Author/Submitter
Take a Test	The Language Learner	MITRE
	Reading Comprehension	MITRE
	The Generic Test Taker	MITRE
	Read a Chapter in a College Freshman Text and Answer the Questions at the End of the Chapter	Raj Reddy
	Build a Large Knowledge Base by Reading Text, Reducing Knowledge Engineering Effort by One Order of Magnitude	Ed Feigenbaum
	Cognitive Decathlon or The Virtual 3rd Grader: California STAR Challenge	Dave Gunning
Analyze and Persuade	The Incident Investigator	MITRE
	The Automated Attorney	MITRE
	The Digital Debater	MITRE
	Handy Andy	Paul Cohen
	Cognitive Decathlon or The Virtual 3rd Grader: Convincing Letter Challenge	Dave Gunning
Learn Then Do / Learn Then Teach	The Device Programmer	MITRE
	The Master Chef	MITRE
	The Tutor and Student	MITRE
	Cognitive Decathlon or The Virtual 3rd Grader: Learning Procedures Challenge	Dave Gunning
	Learn to Read, Read to Learn	Lynette Hirschman
Play a Game	The Multi-Player Strategy Game Challenger	MITRE
	Chess Machine	Raj Reddy
	Learn to Do Crossword Puzzles	Barbara Yoon
Location-Aware Logistical Support	The Digital Dispatcher	MITRE
	The Geo Finder	MITRE
	Ubiquitous Safety.Net	CRA
	Disaster Management	Paul Rosenbloom
	Learn to Use Maps	Barbara Yoon

Personal Assistant	Intelligent Personal Digital Assistant	Bob Balzer
	Context-Aware Information Assistant	Dan Siewiorek
	Memories for Life	UKCRC
	Personal Help Device	Austin Tate
	Lifelong Digital Companion	UKCRC
	Mnemonet	Nigel Shadbolt
	Sensory Augmentation System	Gill Whitney
	Computational Companion for the Old	Yorick Wilks
	Personal Memex	Jim Gray
	Provide a Teacher for Every Learner	
	Reading Tutor	Thomas Kalil
	Employment Support for Disabilities	Thomas Kalil
Scientific Support	Mathematical Discovery	Raj Reddy
	Mathematical Assistant	Toby Walsh
	Automatic Programmer	Jim Gray
	Distilling from the WWW a Huge Knowledge Base, Reducing the Cost of Knowledge Engineering by Many Orders of Magnitude	Ed Feigenbaum
	Medical Safety	Thomas Kalil
Communication Support	The Translating Telephone	Raj Reddy
	Web Understanding Aid	Ehud Reiter
	Learning to Interpret Satellite Images	Barbara Yoon
	Learn a New Language	Barbara Yoon
	Cognitive Decathlon or The Virtual 3rd Grader: Change of Representation Challenge, Book Report Challenge	Dave Gunning
	Speech to Text (Hear as Well as Native Speaker)	Jim Gray
	Text to Speech (Speak as Well as Native Speaker)	Jim Gray
	See as Well as a Person	Jim Gray
Physical Activities	Accident-Avoiding Car	Raj Reddy
	On-Road Driving System	NIST
	Robot Soccer	Manuela Veloso
	Learn to Play Soccer	Barbara Yoon
	Learn to Drive	Barbara Yoon

Collaboration Support	"Smart" Meeting Room Data Collection	NIST
	Build a Team of Your Own	
Creative Activities	Interactive Electronic Musician	David De Roure
	Cognitive Decathlon or The Virtual 3rd Grader: Creative Writing Challenge	Dave Gunning
Question Answering	Deep Thought	Michael Fisher
	Google for Images	Andrew Fitzgibbon, Andrew Zisserman
	World Memex	Jim Gray
Prediction	The Market Predictor	MITRE
Human Impersonation	The Turing Test Game Show Player	MITRE
	Human-Level AI	Raj Reddy
	Model Humans	Paul Rosenbloom
	The Feigenbaum Test	Feigenbaum
	The Turing Test	Alan Turing
	Robot Baby	Paul Cohen
Deception Detection	The Deception Detector	MITRE

Table 2 Previous Grand Challenge Proposals, Categorized

B. Appendix: Sources Consulted for this Review

Author	Description	Notes
UKCRC	submissions to and results from the Grand Challenge development process sponsored by the UK Computing Research Committee (UKCRC), http://www.nesc.ac.uk/esi/events/Grand_Challenges/	100 submissions, approximately, of which approximately 25 were possibly relevant; 7 GCs proposed, of which one was possibly relevant
NIST	a document from Elena Messina at NIST, "Evaluating Cognitive Systems"	list of desiderata for a cognitive challenge problem and for its supporting infrastructure; exemplified through two examples, an on-road driving system and "smart" meeting room data collection
Yoon	five slides from Barbara Yoon (DARPA IPTO) on learning challenges	focuses on learning
CRA	submissions to and results from on a Grand Challenge development conference sponsored by the Computing Research Association (CRA); report @ http://www.cra.org/reports/gc.systems.pdf	70 submissions approximately, of which approximately 8 were possibly relevant; 5 GCs, of which 3 are possibly relevant
Senator	a briefing by Ted Senator (DARPA IPTO) at the Real World Learning Kickoff Workshop, 4/12-13/04	briefing on workshop organization, with one slide (16) on challenge problem criteria
Cohen	slides from Paul Cohen's AAAI talk "If not Turing's test, then what?"	what's right and wrong with the Turing test, and what a good test would look like
MITRE	criteria from MITRE's internal Grand Challenge development exercise for DARPA IPTO; document entitled "'The Grand Challenge' Challenge", October 2003	presents 15 proposed Grand Challenges, broken down by task, technology, and evaluation requirement
Brachman	Ron Brachman, "Systems that Know What They're Doing", IEEE Intelligent Systems, November/December 2002	
Gray	Jim Gray, Microsoft MS-TR-99-50, text of -998 ACM Turing Award lecture "What Next? A Dozen Information-Technology Research Goals", http://research.microsoft.com/scripts/pubs/view.asp?TR_ID=MSR-TR-99-50	presents ~10 GCs, of which 6 are possibly relevant
Feigenbaum	"Some Challenges and Grand Challenges for Computational Intelligence", Edward Feigenbaum, JACM 50.1 (1/2003)	rethinking the Turing Test
Gentner	Gentner, D. (2003). Why we're so smart. In Language in mind: Advances in the study of language and thought (MIT Press). http://www.psych.nwu.edu/psych/people/faculty/gentner/newpdfpapers/GentnerWW03.pdf	essential properties of human cognition

Table 3 Sources Referenced in Analysis